Purification and Biophysical Study of RNase P Proteins 21 and 29 from *Methanocaldococcus jannaschii*

A Senior Honors Thesis by David G. Smith
The Ohio State University
June 2010

Research Advisor: Dr. Mark P. Foster, Department of Biochemistry

ABSTRACT

RNase P is an endonuclease with a rare catalytic RNA core. The enzyme is a ribonucleoprotein (RNP) complex and is primarily responsible for cleaving the 5' RNA leader sequence from precursor tRNA (ptRNA) molecules. The maturation of tRNA is necessary for all domains of life, which makes RNase P an indispensable and ubiquitous enzyme. This historical prevalence renders RNase P as a potential marker for evolution events. Early life forms like bacteria have very few RNase P Protein (RPP) cofactors; whereas later organisms like eukaryotes have been documented as having more than 10 different protein cofactors. The increasing protein content has subsidized, diminishing RNase P RNA (RPR) content. Ongoing studies of RNase P aim to demonstrate how proteins have assumed biological responsibility from nucleic acids.

Our research focuses on the Archaeon *Methanocaldococcus jannaschii* (*Mja*). The *Mja* RNase P has five protein cofactors, two of which are RPP21 and RPP29. The conserved RPP21-RPP29 heterodimer has been implicated in substrate recognition and binding events; however, the latter observation has yet to be applied to Eurkaryotes. *Mja* was chosen for our study because the RPR is very similar to eukaryotic RPR, and may help bridge the evolutionary gaps between bacteria and humans.

Using NMR spectroscopy, our objective is to deduce a solution structure of the RPP21-RPP29 heterodimer. Our efforts have only extended so far as to express and purify both proteins and acquire NMR spectra for RPP29. I have been able to make predictions for the RPP29 secondary structure and have just begun to characterize protein binding; quaternary structure can't yet be determined. These data are a springboard for continuing research that will contribute to the discussion of evolution and protein-protein and protein-RNA interactions.

INTRODUCTION

RNase P Overview

In 1970, Sidney Altman discovered a unique enzyme that contained RNA and protein, but it wasn't until later that decade when serious shock-waves were felt around the scientific community [1]. By 1982, Altman's suspicions were confirmed [2], as RNase P became regarded as an exception to the central dogma of biology: RNA can perform catalysis. RNase P is ribonucleoprotein (RNP) complex which performs the essential step of cleaving of the 5' leader sequence from precursor tRNA (ptRNA) (Fig. 1.1)[3-5]. The RNP complex is found in all domains of life, and each homolog contains one catalytic RNA subunit (RNase P RNA, RPR) and variable numbers of protein subunits (RNase P Proteins, RPPs). In eubacteria, there is only one RPP, and in more highly evolved species—like humans—the RNase P complex contains at least nine protein subunits[6]. Interestingly, amidst the protein diversity, only the conserved RPR core retains catalytic activity [7, 8].

It is certain that archaeal and eukaryotic RPPs have adopted some functions specific to bacterial RPR [9], yet it is unclear why the RNA still remains. *In vitro* experiments have demonstrated that the RPR alone is sufficient for catalysis; however, catalytic activity in eukaryotic and some archaeal RPR can only be recovered under specific circumstances [10, 11], but even then, they have a slower turnover than bacterial homologs. *In vivo*, the RPPs are essential for catalysis [12]. Whether they mitigate some thermodynamic barrier or support the RNA tertiary structure [13-15], their role as a cofactor belies their vastly greater modularity and diversity over RNA. Adding more proteins to the RNP seems averse to Occam's razor. Our interest is in resolving why more proteins are found in more recently evolved species and discovering if there is evidence for diverging RPR structures being superseded for more modern protein-based biology.

Scrutiny of RNase P has revealed the molecule's breadth of activity. Besides processing tRNA, RNase P processes a variety of other ptRNA-like substrates [15-19] and regulates gene transcription [20]. In regard to human health, RNase P is an attractive antibiotic target because any disruption of the

RNP complex inhibits enzymatic function [21].

Bacterial RNase P

Bacterial RNase P is the simplest RNP configuration, one RPP cofactor and one large RPR [22]. The bacterial RPR can be classified into two types, type A (e.g. *E. coli*) and type B (e.g. *B. subtillis*) (Fig. 1.2) [22]. The RPR structures for both bacterial types have been reported by x-ray crystallography [23-26], and the RPP structures have been reported by crystallography and NMR [27, 28], but a high resolution structure of the holoenzyme has not yet been achieved.

Being the best studied, several general characteristics of RNase P have been extrapolated from bacterial forms. Namely, the RPR is composed of two domains, the C-domain and S-domain. The catalytic core is located on the C-domain, and when folded, it forms important intramolecular contacts with the S-domain [28]. The ptRNA cleavage reaction is performed by Mg^{2+}-coordinated H_2O molecule via a nucleophilic attack on a scissile phosodiester bond [29]. During this process, the conserved T stem loop of tRNA interacts the RPR S-domain [30], the tRNA acceptor stem interacts with the C-domain [31], and 3'-CCA (tRNA) base-pairs to the L15 region [32] (Fig.1.2).

RNase P in higher organisms

In Archaea, RNase P is composed of at least four RPPs and one RPR, and like bacterial RNase P, the archaeal RPR is categorized in two distinct groups, type A (*P. furiosus, Pfu*) and type M (*M. jannaschii, Mja*)(Fig 1.2) [33, 34]. Type A closely resembles bacterial RPR, and type-M closely resembles eukaryotic RPR, from which it differs by two additional RNA regions. All RNRs from higher organisms are distinguishable from bacterial RNR by their deleted sequences [33]. Furthermore, the pattern of deleted RNA structural elements is consistent among archaeal and eukaryotic RNR types. It's reasonable to propose that proteins were selected to replace lost RNA structure.

Neither archaeal nor eukaryotic RPPs share sequence similarity to the bacterial RPP, however, evidence suggests that four of the archaeal RPPs share sequence similarity with human RPPs [33, 35]. The four archaeal proteins (RPP21-RPP29 and POP5-RPP30) form binary complexes prior to binding

to the RNA [36], and each pair exerts some kinetic benefit upon the RNR. The structures of each of these four archaeal RPPs have been solved from various archaea with either X-ray crystallography or NMR. Not surprisingly, each structure contains familiar nucleic acid binding motifs: RPP21 contains a Zn (zinc) ribbon [37], RPP29 has an Sm-like fold (Sm protein) [38], RPP30 has a TIM-barrel (Triose-phosphate isomerase) [37], and POP5 has an RRM-like fold (RNA recognition motif) [39].

Unlike the bacterial RPRs, in the absence of the proteins, M-type RPR can only perform catalysis under high ionic conditions and when the substrate is tethered in cis (Fig. 1.3) [11]. Reconstitution assays have shown that either binary complex is sufficient for archaeal RPR to process ptRNA, but no one RPP can elicit product turnover (Fig 1.4) [40]. Most notably, the POP5-RPP30 complex significantly increases the K_{conf}, which is the equilibrium constant for the transition state conformations of RNase P-ptRNA. A similar rate enhancement is observed from RPP21-RPP29, but the effect is smaller. It is uncertain whether the eukaryotic RPP homologs have similar effects [41, 42].

Archaea are valuable model organisms because they provide a manageable system for understanding their eukaryotic homologs, which are often too complicated for detailed study. Extrapolations from bacterial RNase P may be less reliable because of the substantial differences between bacterial and eukaryal RNA and protein sequences. A structure for the free RPP21-RPP29 complex has been published for *Pfu* [43], and if a structure could be solved for *Mja*, then it would be possible to generate testable hypotheses for how proteins have compensated for diminished RNA function (Fig 1.7).

Methanocaldococcus Jannaschii

Interest in *Mja* RNase P stems from its unique RNR similarity to eukaryotic RNR. Additionally, there are similarities between RPPs (RPP21 is 34% similar to human homolgs and RPP29 is 20% similar to human homologs), though comprehensive comparisons to eurkayotic homologs are severely handicapped by insufficient data and poorly understood RNase P mechanisms.

The divergent evolution of the RPR into type-M and type A is best characterized by the loss of

structure. Archaeal type-A RPR is similar to bacterial ancestors but is distinguished by the absence of P18, P13, and P14 regions (Fig. 1.2). Type A has a larger, conserved P12 structure and is identified as the ancestral structure class for type-M RPR. Type M is distinguished by a significant rearrangement of the cruciform (P7-P11) and the absence of everything distal to P15. A consequence of these deletions is the disappearance of P6, and the disappearance of RNA structures responsible for substrate recognition (Fig. 1.2) [34]. Consistent with RNA world predictions [44], type M RNA contains no additional RNA structural elements that could compensate for the deleted RNA.

Despite distinct intramolecular interactions and RNA activity between the domains of life, there seems to be no correlation between the number of RPPs and the RPR type. This allows the possibility for protein structure to function examination. Currently, the only structural data that exists for M-type RPP21-RPP29 comes from homology modeling so our research is aimed at generating a high resolution model of this pivotal complex, which likely represents a discrete step toward modern protein based biology.

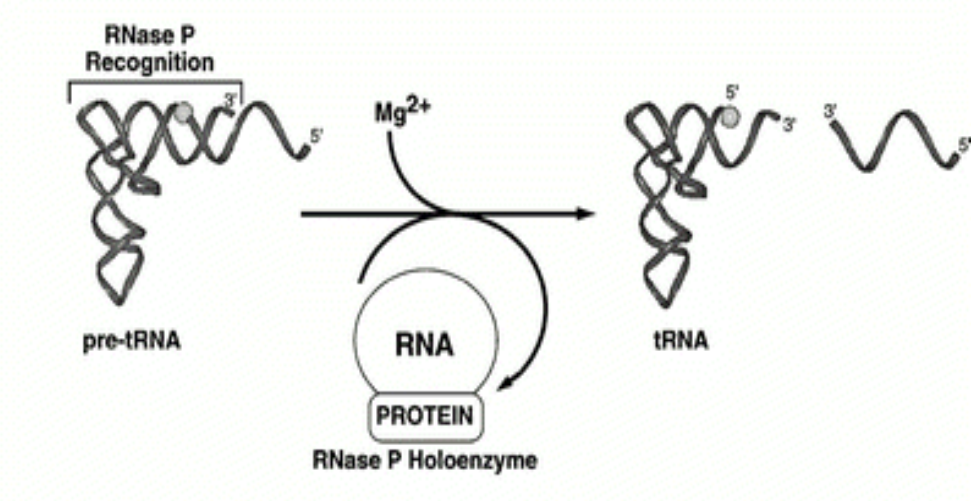


Figure 1.1. The RNase P ribozyme catalyzes hydrolysis of the 5' RNA leader sequence from precursor tRNA. The mechanism requires divalent ions and a protein cofactor. The protein cofactors vary in number and function across the domains of life; however the RNA subunit is thought to be responsible for directly mediating hydrolysis. Figure adapted from ref [9].

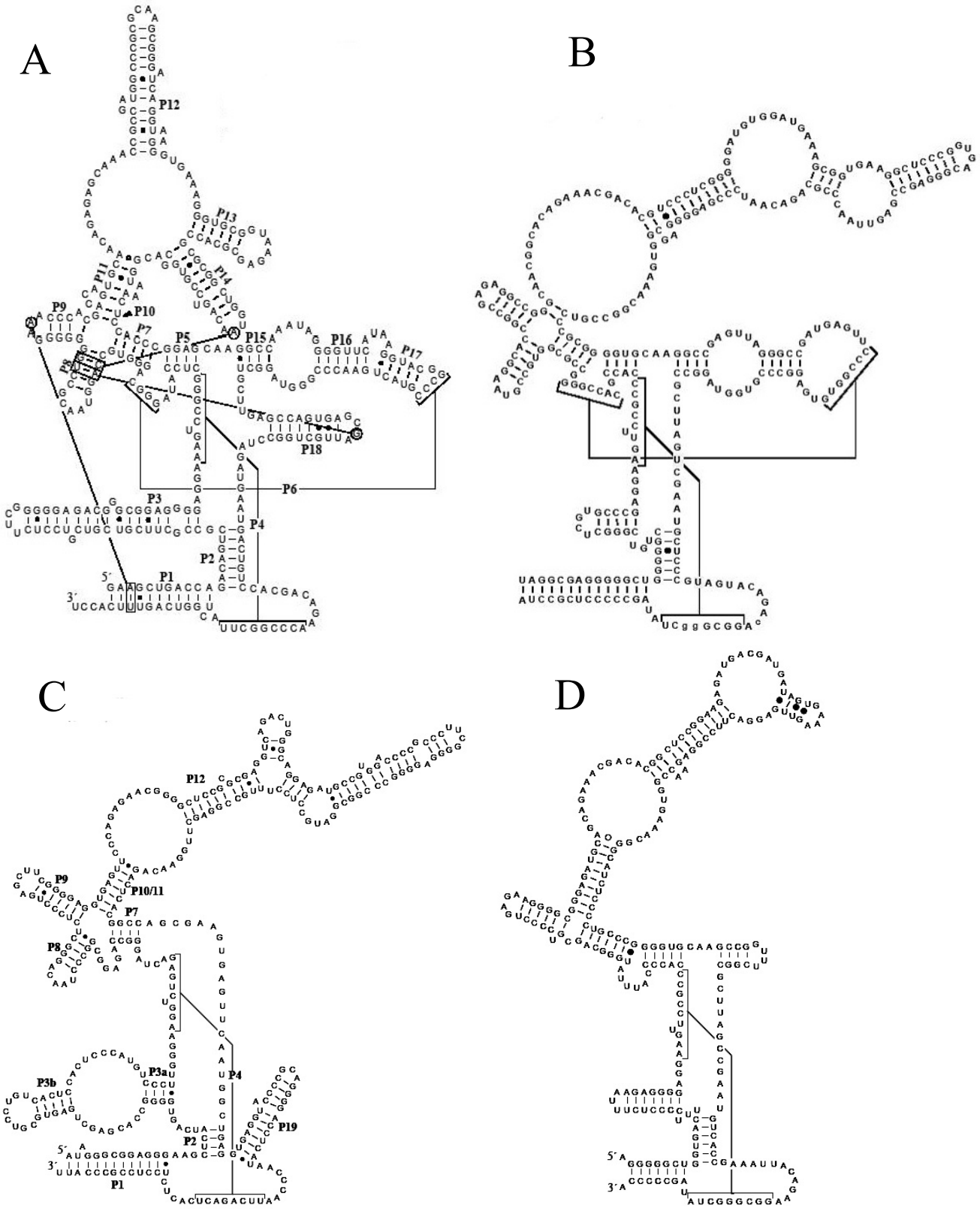


Figure 1.2. The RNase P RNA has unique and conserved features across the domains of life. Helices are designated P1-18. (A) The predicted RPR from Bacteria (*E. coli* is shown) is the largest and most complicated RNA subunit. (B) The predicted secondary structure for archaeal type A RPR retains many of the structural elements found in Bacteria, but some intramolecular contacts have been lost. (C) The predicted secondary structure for eukaryotic RPR bears structural features similar to archaeal type-M orthologs. (D) Predicted archaeal type M RPR secondary structure is simplified from ancestral forms by the loss of P16/P17 regions and a diminished cruciform structure. Figure adapted from ref [45].

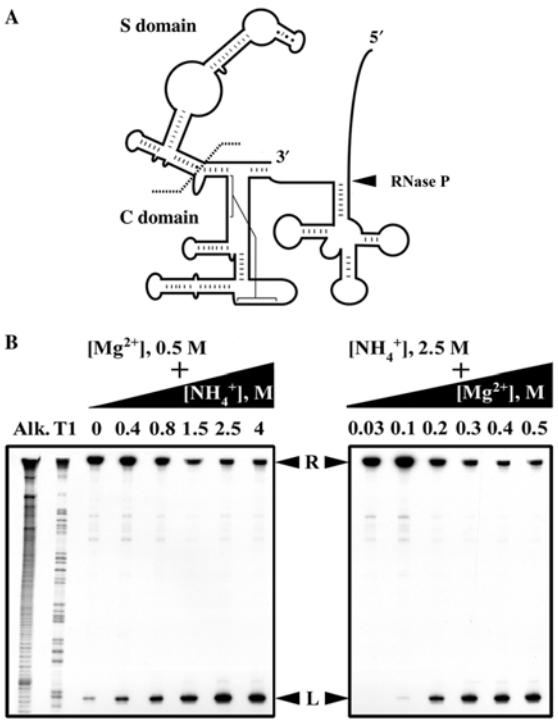


Figure 1.3. *In vitro* assays have been performed that demonstrate RPR alone catalysis. (A) As opposed to archaeal type-A, type-M RPR is only capable of cleaving ptRNA substrates if they are covalently attached. (B) The extent of RPR-ptRNA self-cleavage increases as ion concentration increases. The band labeled L represents the 5' ptRNA leader sequence, and the band labeled R represents the tRNA and RNR. Figure adapted from ref [11].

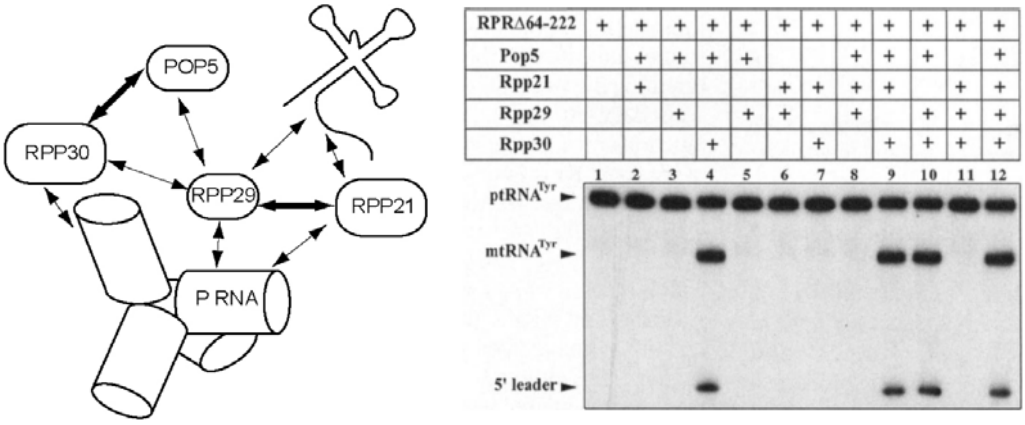


Figure 1.4. (Left) The RNP complex is composed from an intricate array of protein-protein, protein-RNA, and RNA-RNA interactions. (Right) An experiment was performed with *Pfu* that demonstrates how various RNP compositions affect the ribozyme activity. A (+) indicates that the particular RPP was present and a blank indicates the absence of a particular RPP. RPRΔ64-222 is the C domain of the wt RNR. Figure adapted from ref [40].

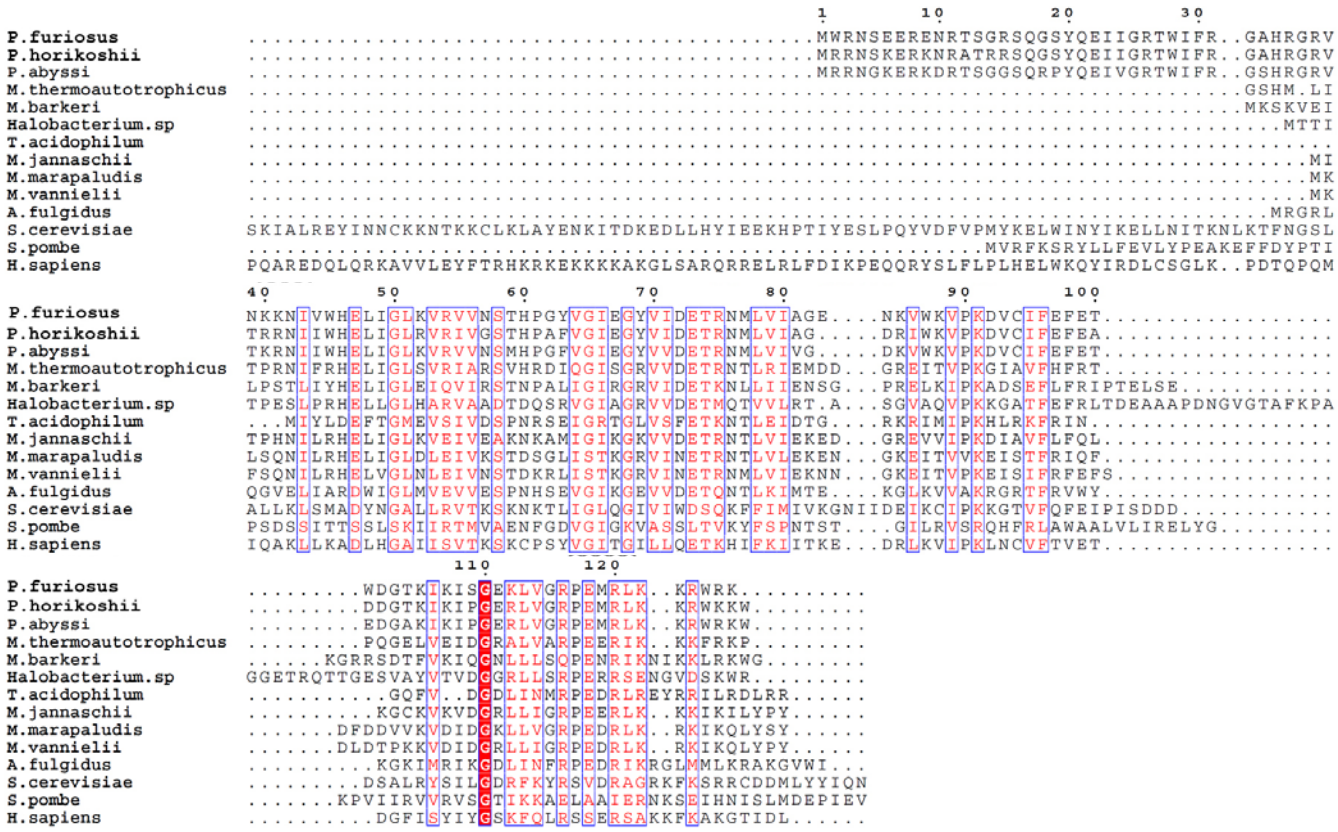


Figure 1.5. Sequence alignment for RPP29 from various organisms. Red columns indicate a protein residue invariant in each of organisms, and outlined columns represent a conservation of residues character (e.g. polar, non-polar, acidic, or hydrophobic). Aligned sequences are form *Pyrococcus furiosus* (NCBI entry NP_579545), *Pyrococcus horikoshii* (NP_143607), *Pyrococcus abyssi* (NP_126024), *Methanobacterium thermoautotrophicum* (10QK_A), *Methanosarcina barkeri* (YP_303669), *Halobacterium sp.* (NP_280464), *Thermoplasma acidophilum* (NP_394719), *Methanococcus jannaschii* (NP_247439), *Methanococcus marapaludis* (YP_001549311), *Methanococcus vannielii* (YP_001323236), *Archaeoglobus fulgidus* (1TSF_A), *Saccharomyces cerevisiae* (NP_009816), *Schizosaccharomyces pombe* (NP_588479) and *Homo sapiens* (NP_006618).

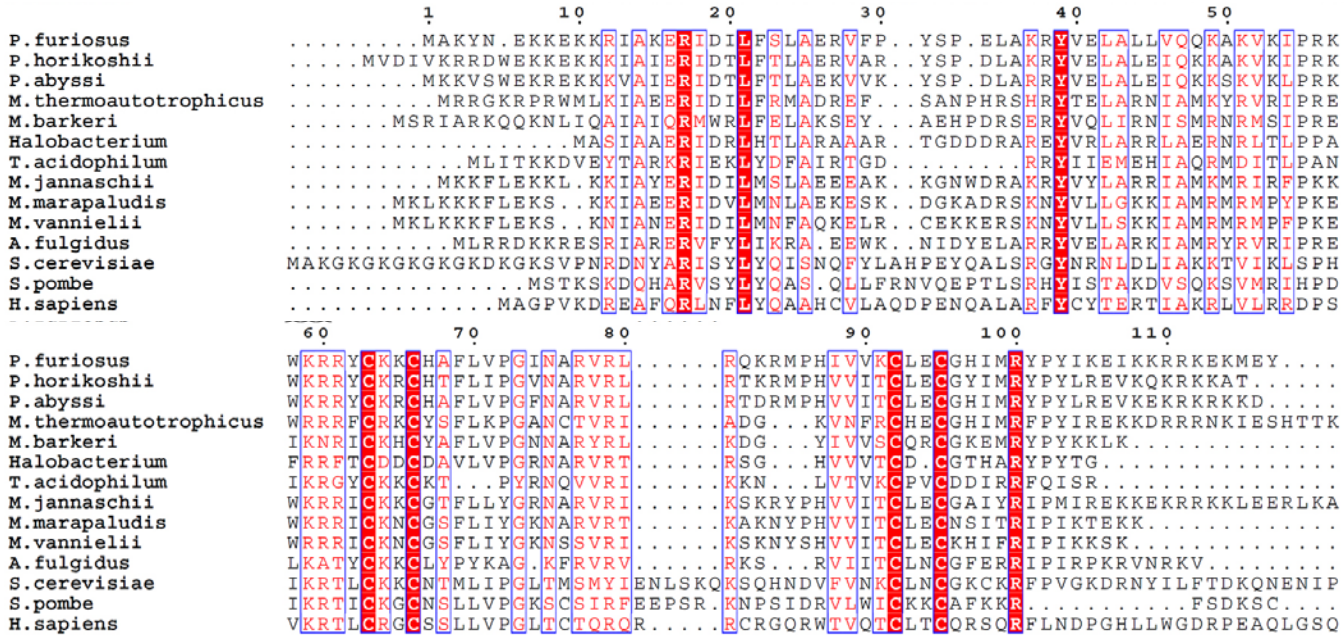


Figure 1.6. Sequence alignment of RPP21 from various organisms. Red columns indicate a protein residue conserved in each of organisms, and outlined columns represent a conservation of residues character (e.g. polar, non-polar, acidic, or hydrophobic). The characters that appear above the sequences refer to likely secondary structures. Aligned sequences are form *Pyrococcus furiosus* (NCBI entry NP_579342), *Pyrococcus horikoshii* (NP_143456), *Pyrococcus abyssi* (NP_126253), *Methanobacterium thermoautotrophicum* (NP_276730), *Methanosarcina barkeri* (NCBI_entry YP_304815), *Halobacterium sp.* (NP_279631), *Thermoplasma acidophilum* (NP_393654), *Methanococcus jannaschii* (NP_247957), *Methanococcus marapaludis* (YP_001549778), *Methanococcus vannielii* (YP_001322736), *Archaeoglobus fulgidus* (NP_068950), *Saccharomyces cerevisiae* (NP_012280), *Schizosaccharomyces pombe* (NP_596472) and *Homo sapiens* (NP_079115).

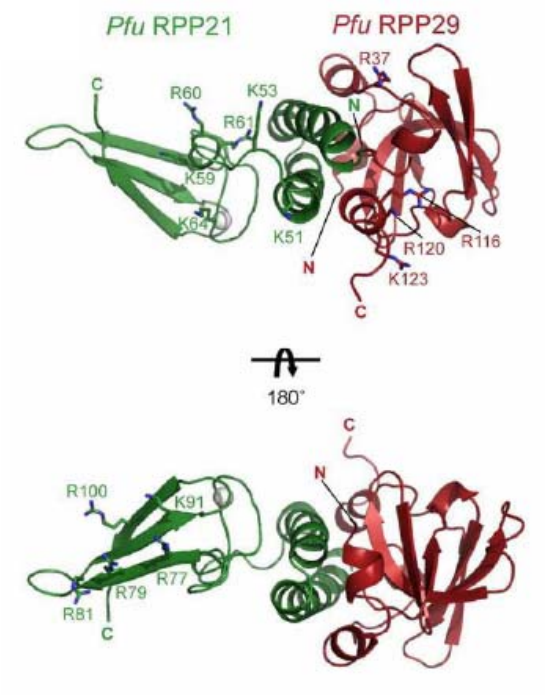


Figure 1.7. Ribbon diagram of the RPP21-RPP29 complex from *Pfu*. The alpha helices of both proteins compose the dimer interface. The structure was solved with NMR. Figure adapted from ref [43].

NMR Spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is the primary biophysical method employed in this study. As opposed to X-ray crystallography, NMR is preferred for several reasons. 1) NMR is obtained from molecules in solution, as opposed to in crystals, where the conformation may deviate from the molecule's biologically relevant form. 2) NMR can provide information about molecular dynamics. 3) NMR can be used to observe molecular complexes with weak binding affinities, which may not be able to form when crystallized due to conformational bias of crystallization conditions.

The theory behind NMR can be interpreted with either classical or quantum mechanics. In quantum mechanics, every particle can be described by a set of quantum numbers. The nuclear spin quantum number is especially relevant for NMR. In principle, each atom in a molecule has a resonance energy and can be excited with the use of electromagnetic energy [46]. The resonance energy of each nucleus can be identified, and its position with regard to nearby atoms can be extrapolated (Fig. 1.9).

To deduce a structure from NMR spectroscopy many different experiments are performed. The first step in a typical procedure is to assign a resonance frequency to each of the ^{1}H, ^{13}C and ^{15}N atoms of the protein. Next, a series of NMR experiments are performed to determine how far each proton is from other protons in the molecule. Then, using distance and torsion angle restraints, a computer program can calculate a set of structures consistent with the input constraints.

1D spectra: Because resonance frequencies for different types of nuclei are very different, a 1-dimensional spectrum only reports the resonance energies (frequencies) of one type of nucleus; for proteins typically ^{1}H (Fig 1.9). Though simple, signal dispersion is a useful metric to determine how well a protein is folded. From a 1D spectrum, many important experimental variables can be optimized: protein concentration, temperature, buffer conditions, and sample purity.

Heteronuclear Single Quantum Coherence (HSQC): An HSQC reveals the through-bond correlation between different nuclei, like ^{1}H and ^{15}N. This experiment is very versatile, as each type of nucleus (e.g. C, N, or H) can be selectively excited, and the transfer of resonance frequency can be easily

controlled (Fig 1.10). HSQC and similar techniques allow researchers to determine the resonance assignment for each atom along a peptide backbone.

Heteronuclear NOE: The Nuclear Overhauser Effect (NOE) involves the through-space transfer of magnetization from one nucleus to another. The magnitude of the NOE depends on the distance between cross-relaxing nuclei, and on the timescale of their motion. The goal of the heteronuclear NOE experiment is to measure how two bonded atoms, with fixed internuclear distances, are moving compared to others. The data collected from the experiment are presented as the ratio between the intensities of signals recorded in the absence and presence of the NOE, and ratio's magnitude reflects how fast a particular pair of nuclei is moving compared to the natural tumbling of the protein (Fig. 1.8).

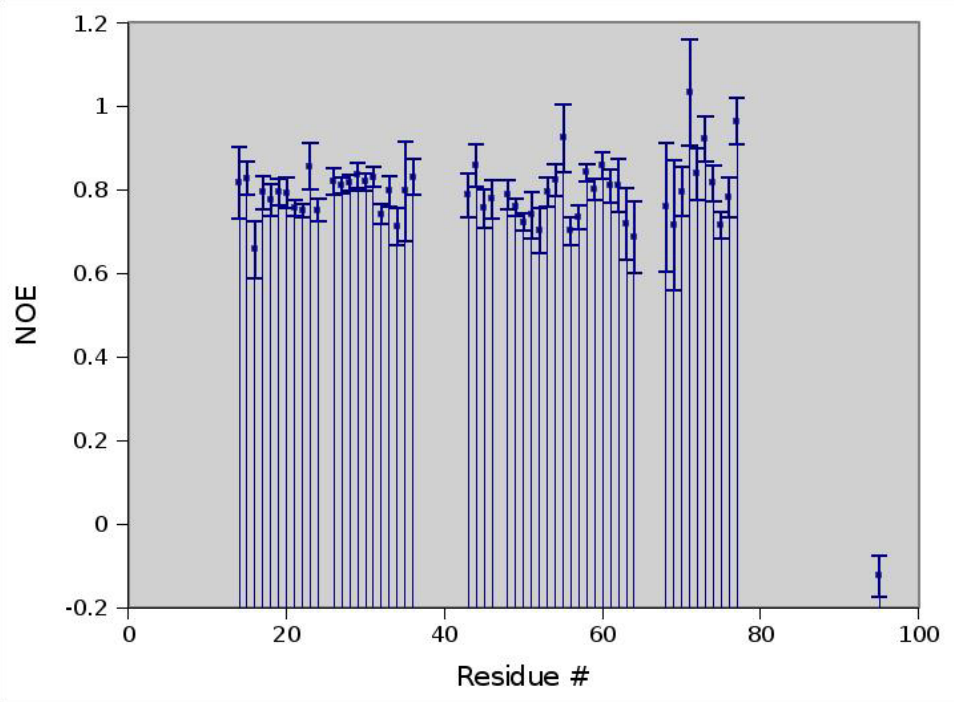


Fig 1.8. Heteronuclear ^{1}H-^{15}N Heteronuclear NOE values for backbone amides in a protein. The y-axis is a measurement of the NOE, and reflects the degree of order. Lower NOE values indicate more motion on the ps-ns timescale.

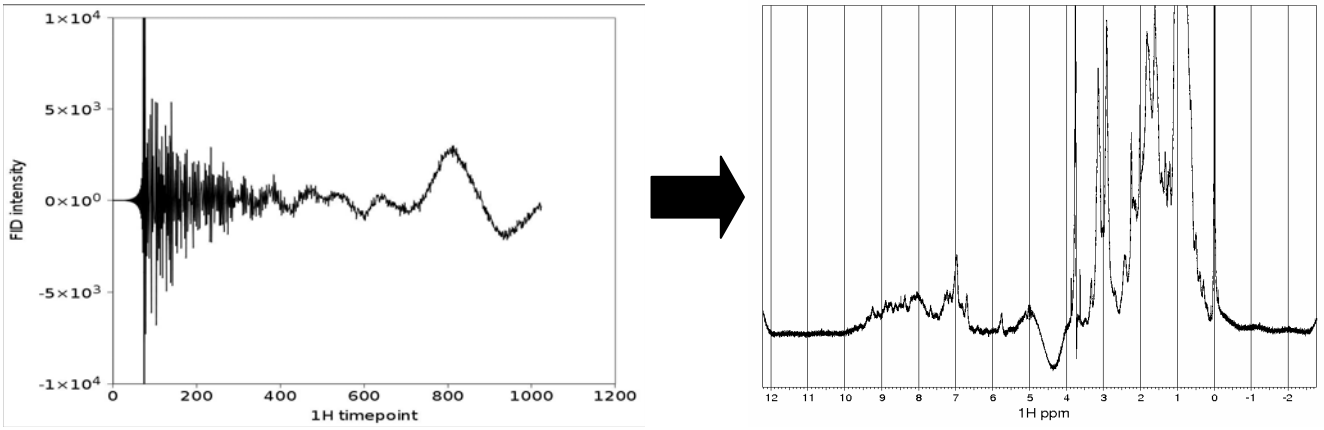


Figure 1.9. An NMR spectrometer measures an oscillating current that decreases with time. (A) FID is the Free Induction Decay of transverse magnetization (B) The FID is reinterpreted via a Fourier transform to give a series of proton signals. The amide protons in a protein resonate in the 6-10 ppm range.

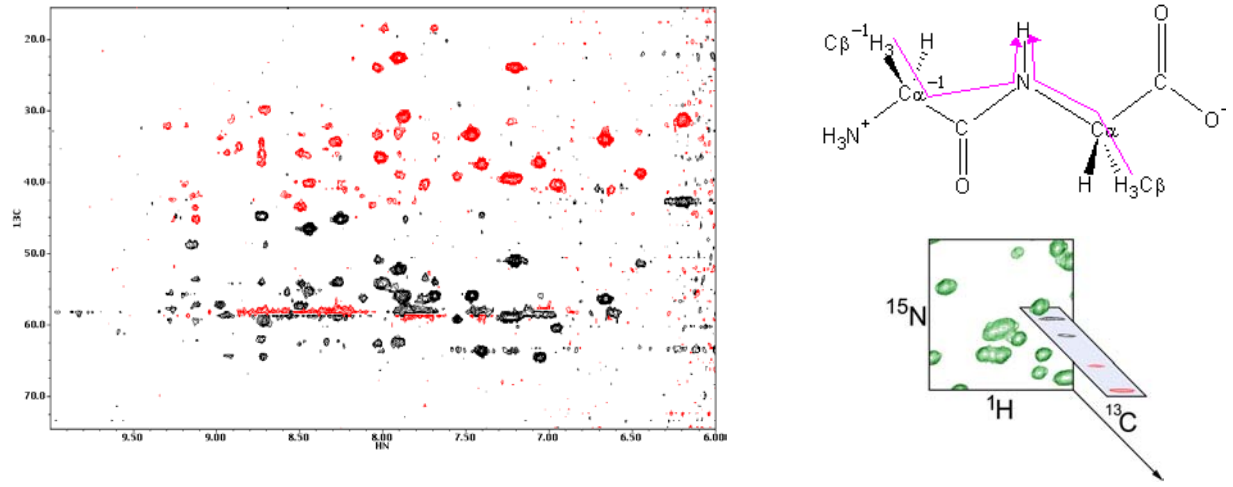


Figure 1.10. Correlation NMR experiments are a method for viewing correlations between different nuclei in the protein. (Left) An HNCACB experiment displays peaks for ^{13}C nuclei that are correlated to nearby protons. (Right) The pink arrows on the molecule show how magnetization is transferred between the protein backbone and sidechain. The bottom figure shows how an HNCACB spectrum is arranged with an ^{15}N-^{1}H HSQC to create a three dimensional graphic.

MATERIALS AND METHODS

Protein Expression

Each of the proteins, RPP21 and RPP29, were independently over-expressed in competent *E. coli* cells. The RPP21 construct was cloned into the Nco1 and BamH1 restriction sites a pET-15b plasmid (provided by J.W. Brown at NCSU), and the RPP29 construct was cloned into the Nde1 and BamH1 restriction sites of a pLANT-2b plasmid (provided by J.W. Brown at NCSU, kanamycin resistant) (Fig. 2.1). Plasmid transcription is controlled by a lactose, negative inducible, operon. The pET-15b plasmid was transformed into Rosetta DE3 (chloramphenicol and carbenicillin resistant) via electroporation, and the pLANT-2b plasmid was transformed into BL21-DE3 (chloramphenicol resistant, rare tRNAs) via electroporation. The bacteria were inoculated onto 3:5, agar (15g/L): Lysogeny Broth (LB) with the appropriate antibiotics (30 mg/L kanamycin, 50 mg/L carbenicillin, 34 mg/L chlormaphenicol). The LB culture plates were incubated for 10-12 hours at 35ºC, after which a single colony was inoculated into 100 mL of LB with the appropriate antibiotics. The 100 mL culture was incubated for 10-12 hours at 35ºC. Afterward, 10 mL of this overnight culture was inoculated into 1 L of LB. The 1 L culture was grown until $0.6<OD_{280}<0.8$. While in this target range the bacteria were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The Rosetta cells also required 50 μM $ZnCl_2$ for adequate overexpression of RPP21, a zinc-binding protein. The cultures were allowed to grow for 10-12 hrs before centrifugation (JS-4.2 rotor, 4,500 *g*, 30 min). The pelleted cells were stored at -4ºC. All LB mediums were autoclaved before use.

Expression of Isotopically Labeled RPP29

When growing ^{15}N, ^{13}C-labeled RPP29, the 1 L LB medium was replaced with minimal media. Preparation of M9 minimal media: 200 mL of 50 mM Na_2HPO_4, 25 mM KH_2PO_4, 10 mM NaCl, 20 mM $^{15}NH_4Cl$ combined with 750 mL of 1M $MgSO_4$, 20 mM ^{13}C-dextrose, 1M $CaCl_2$. The growth

medium was autoclaved, then 10 mL of Gibco Basal Eagle Vitamin mix and antibiotics (1 mM) were added. 10 mL of bacteria culture (OD_{600} = 1.0) were inoculated into the M9 minimal media and the procedure for protein expression was followed.

Protein Purification and Chromatography

RPP29: From a 1 L culture, pelleted cells were resuspended in 25 mL of buffered 25 mM Tris, pH 7.5, 25 mM NaCl, 0.1 mM PMSF, 1 mM EDTA. The cells were lysed on ice via sonication (5 min, 5 sec pusles, 2 sec int., 65 W)x2. After centrifugation (SS-34 rotor, 26,892 *g*, 15 min), the supernatant was decanted and the pellet was resuspended in 25 mL of buffered 7 M urea, 25 mM Tris, pH 7.5, 10 mM DTT, 1 mM EDTA (resuspension buffer). The resuspended lysate was centrifuged (SS-34 rotor, 26,892 *g*, 15 min) and filtered (0.1 micron), then loaded onto a 5 mL cation exchange column equilibrated with 25 mL of resuspension buffer. The protein was eluted using fast protein liquid chromatography (FPLC) with an eluent gradient of 0 mM NaCl to 2M NaCl. The eluate fractions containing RPP29 were pooled and extensively dialyzed against buffered 10 mM Tris, pH 6.7, 10 mM NaCl, 0.02% NaN_3, 0.03 mM $ZnCl_2$ (NMR buffer). The protein was concentrated down to 0.5-0.75 mL and stored at -4ºC.

RPP21: From a 1 L culture, pelleted cells were resuspended in 25 mL of buffered 25 mM Tris, pH 7.5, 25 mM Kcl, 5 mM imidazole, 6M guanidinium-HCl (lysis buffer) and lysed on ice via sonication [5 min, 5 sec pusles, 2 sec int., 65 W]x2. The lysate was centrifuged [SS-34 rotor, 26,892 *g*, 15 min], and the supernatant was filtered (0.1 micron) and loaded onto 5 mL metal chelating column (Ni^{2+}) equilibrated with 25 mL of lysis buffer. The protein was eluted using FPLC with a gradient from 50 mM imidazole to 0.5 M imidazole. The eluate fractions were pooled and extensively dialyzed against buffered 10 mM Tris, pH 6.7, 10 mM NaCl. The $(His)_6$ tag was removed via TEV protease [1:30, RPP21:TEV]. Optimal cleavage was achieved at room temperature in 10 mM Tris, pH 6.7, 10 mM NaCl. The TEV, $(His)_6$ tag, and uncleaved protein were separated with a second Ni^{2+} column. RPP21

was washed off the column with 10 mL of buffered 25 mM Tris, pH 7.5, 25 mM KCl, 50 mM imidazole, 8M urea, and extensively dialyzed against (10 mM Tris, pH 6.7, 10 mM NaCl, 0.02% NaN_3, 0.03 mM $ZnCl_2$). The purified protein was concentrated down to 0.5-0.75 mL and stored at -4ºC.

NMR Spectroscopy

Sample Preparation

To prepare sample of the RPP21-RPP29 complex in which RPP29 was uniformly labeled, the proteins were dialyzed into the same buffer (10 mM Tris, pH 6.7, 10 mM NaCl, 0.02% NaN_3, 0.03 mM $ZnCl_2$) and the unlabeled protein (RPP21) was prepared in excess of the labeled protein (RPP29) (1.2:1). DSS was then added to the sample (5x[protein]) and the proton concentration was diluted with D_2O (10% *v/v*). The NMR sample was prepared at 0.6-0.8 mM.

Data Acquisition

All NMR spectra were acquired from a 600-MHz Bruker Avance DRX spectrometer equipped with cryogenically cooled triple-resonance pulse-field gradient probes. Preliminary 1D and 2D (^{15}N-^{1}H) spectra were recorded at 25ºC, 37ºC, and 55ºC. Subsequent spectra were all acquired at 55ºC. The ^{1}H, ^{15}N, ^{13}C resonance assignments for free RPP29 were obtained from data collected in the following experiments: ^{15}N-^{1}H HSQC, HNCO, HNCA, CBCA(CO)NH, and HNCACB [46]. Another ^{15}N-^{1}H HSQC was collected from a sample containing [^{15}N] RPP29 and unlabeled RPP21. The last experiment performed was a heteronuclear {^{1}H}-^{15}N NOE on free RPP29. All NMR spectra were processed and analyzed with NMRPipe [47], NMRView [48], and CARA (Computer Aided Resonance Assignment [49]).

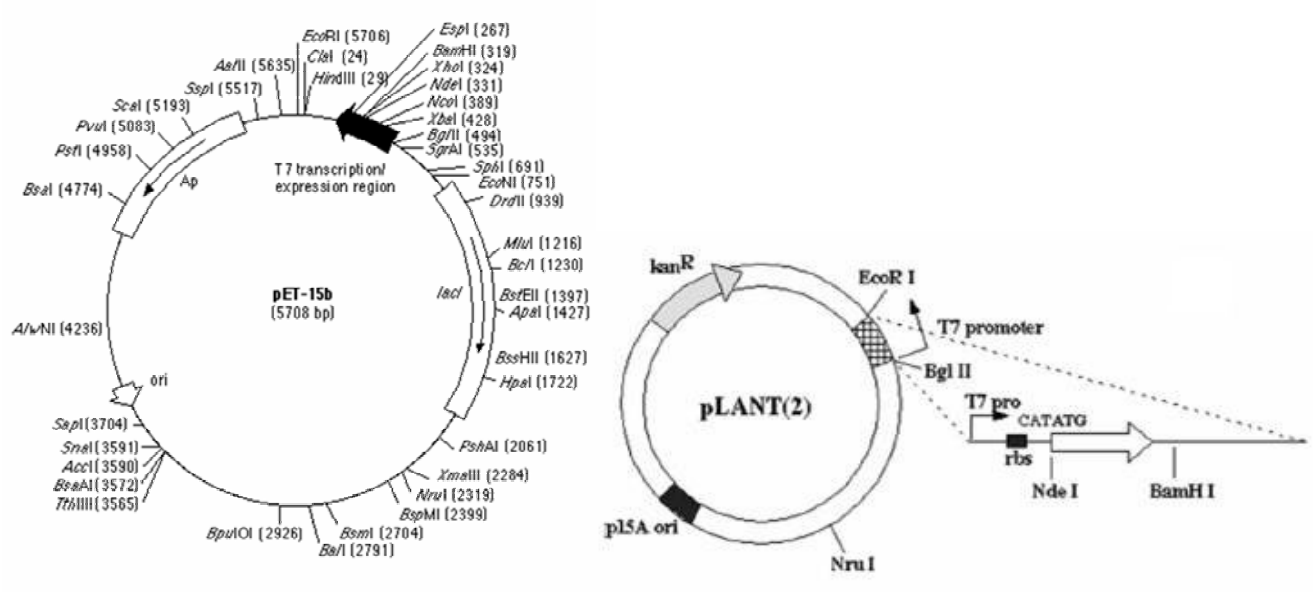


Figure 2.1. (Left) Map of the pET-15b plasmid vector [50]. *Mja* RPP21 was cloned into sites Nco1 (5') and BamH1 (3') (Right) pLANT-2b plasmid vector [51]. *Mja* RPP29 was cloned into sites Nde1 (5') and BamH1 (3') [8].

RESULTS

RPP29 is over-expressed in *E. coli*

The pLANT-2b plasmid that was transformed and expressed in BL21(DE3) cells yields 30-50 mgs of RPP29 per liter of LB. The produced protein was concentrated in the insoluble fraction of the cell lysate, perhaps due to the formation of inclusion bodies, and was resuspended in denaturing conditions (7M urea). Based on the pIs of constituent amino acids, the unfolded protein has a predicted pI of 9.6 so at pH 7, the majority of RPP29 in solution is expected to be cationic. A 5 mL SP Sepharose cation-exchange column (GE Healthcare) was used to isolate the protein, which eluted in 200 mM NaCl. RPP29 was easily resolved from most other protein eluates. When refolding the protein by dialysis into buffer without urea, white precipitation was observed. SDS-PAGE confirmed the precipitate to be contaminating protein. The precipitate was separated via centrifugation and pure RPP29 was obtained (Fig. 3.1).

RPP21-$(His)_6$ is over-expressed in *E. coli*

The pET-15b plasmid encoding the His_6-RPP21 fusion construct with a TEV cleavage site (…HH-NLYFQ/G-RPP21) was transformed and expressed in Rosetta, yielding 60-80 mgs of RPP21 per liter of LB culture. The protein eluted from the Ni^{2+} column (GE Healthcare) in 125 mM imidazole. The protein was refolded via dialysis into non-denaturing buffer; however, Zn^{+} was excluded from the buffer because it interferes with the TEV protease. The $(His)_6$ tag was successfully removed after 24 hrs of digestion (Fig. 3.1). The Ni^{2+} chelating resin binds histidine so cleaved RPP21 poorly binds to the column and could be obtained after washing the column. The isolated and cleaved protein contained an additional glycine residue at the N-terminus, but it proved to be inconsequential during dialysis refolding.

Purification of RPP21 and RPP29

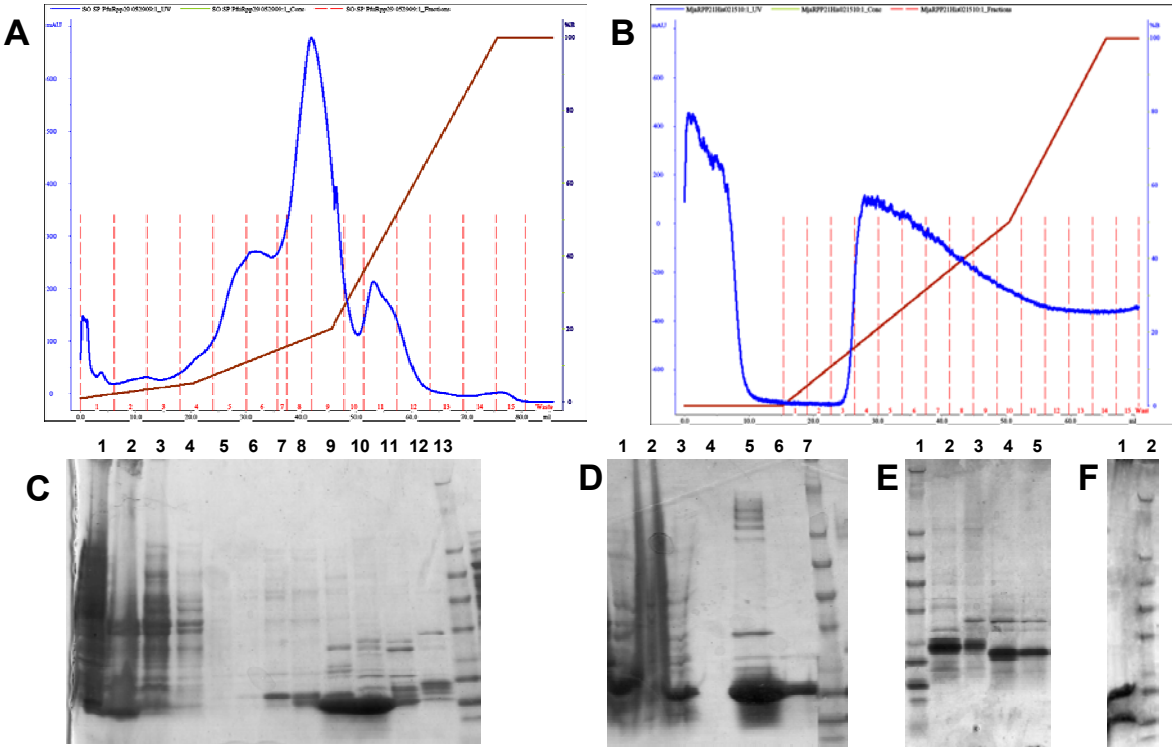


Figure 3.1. (A) The red line on the chromatogram represents the concentration of NaCl in the column. When the concentration of salt reaches 0.2 M (10% buffer B), RPP29 begins to elute (the largest peak). The blue line is the UV absorbance of eluting material and the dashed red lines are eluate fractions. Beneath the chromatogram (C) is an SDS-PAGE gel showing the stages of RPP29 purification. Lanes: (1) soluble fraction after lysis (2) SP column load-on (3) column flow through (4) column wash (5-12) column eluate fractions 4-11 (13) molecular weight ladder. (B) The red line represents the concentration of Imidazole in the Ni^{2+} column. The blue line represents the UV absorbance of eluting material and the dashed red lines are the eluate fractions. Beneath the chromatogram are SDS-PAGE gels. (D) Stages of RPP21 purification. 1) column load on 2) column flow through 3) column wash 4)-BLANK- 5) protein eluate #3 6) protein eluate #11 7) molecular weight ladder. (E) TEV cleavage of His tag 1) molecular weight ladder 2) RPP21-$(His)_6$ 3) RPP21-$(His)_6$ w/ TEV in Zn^{2+} 4) RPP21-$(His)_6$ w/ TEV no Zn^{2+}, 4°C 5) RPP21-$(His)_6$ w/ TEV no Zn^{2+}, 25°C. (F) Purified protein 1) RPP21/RPP29 2) molecular weight ladder

NMR spectroscopy of free RPP29

Two-dimensional ^{1}H-^{15}N correlated NMR spectra were acquired from unbound RPP29 at three temperatures; 25ºC, 37ºC, and 55ºC (Fig. 3.2). At 25ºC, the spectrum linewidths were very broad and few resolved peaks could be discerned from the spectrum. Additionally, the signal to noise ratio (S/N) was worse than that observed in other spectra (37ºC and 55ºC). The poor quality of the spectrum was likely due to semi-aggregation. At 37ºC, the linewidth had improved, and more peaks could be resolved. The highest quality spectrum was obtained at 55ºC; with 130 identifiable spin systems and the narrowest linewidths. The S/N was also vastly improved. The triple resonance experiments were performed next. The NMR spectra (HNCO, HNCA, CBCA(CO)NH, and HNCACB) were recorded over a 2 day period.

CARA was used to make the resonance assignments (Fig. 3.3). Of the 95 amino acids that compose RPP29, 55 could be assigned (residues 14-24, 26-36, 43-46, 48-54, 56-65, 68-77, and 95). With the exception of the C-terminal tyrosine, all of the identified backbone resonances are located within the core sequence of the protein. The Sm-like fold expected for RPP29 is characterized by a series of anti-parallel beta strands and alpha helices at both termini. The C-terminal alpha helix is expected to form between residues 74-86; however, no chemical shifts were observed from that region. This suggests that the alpha helix may only form upon binding of RNA or RPP21. Similarly, the absence of identifiable signals from the N-terminus is indicative of ill-defined structure for that region of the protein.

The $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N chemical shifts from assigned backbone resonances were compared to chemical shifts expected from random coil peptides. Similar and consecutive deviations from the expected random coil value were interpreted as secondary structure. The value of the deviation was used to predict whether each residue adopts an alpha helical or beta sheet structure. Consistent with an

Sm-like fold and homology modeling, the data predicted a beta sheet, followed by an alpha helix, then four more beta sheets (Fig. 3.4).

The $\{^{1}H\}$-^{15}N NOE revealed that the core residues of unbound RPP29 were rigidly structured. NOE values close to 1 imply a high degree of structure, whereas smaller or negative values imply a high degree of flexibility. Most of the assigned resonances had a value of 0.8. The only residue with a distinctly lower NOE value was tyrosine-95, which should be expected from a highly disordered area (Fig. 1.8).

Spectral perturbations of RPP29 in solution with RPP21

Acquiring NMR spectra from isotope labeled RPP29 in solution with unlabeled RPP21 proved to be troublesome. Repeated experiments produced similar spectra with irregular lineshapes and a noise ridge in the ^{1}H domain (Fig 3.5). I was not able to determine where or how well RPP21 bound to RPP29, but there are clear perturbations between the bound and unbound RPP29 spectra. The most Intriguing perturbations occurred within the β5 strand (residues 68-74), which lost most of the unbound amide signals (Fig 3.5). This region was the prominent exception, as signals appeared for nearly every assigned resonance. It was common to observe an amide signal overlapping an unbound peak while another peak was observed in the adjacent area (Fig. 3.6). This observation is most likely due to an inhomogeneous mixture of RPP29 folding conformations; unbound and otherwise. Apart from residues 68-72 and 74, only three other assigned resonances were conspicuously changed, while at least five new signals were observed.

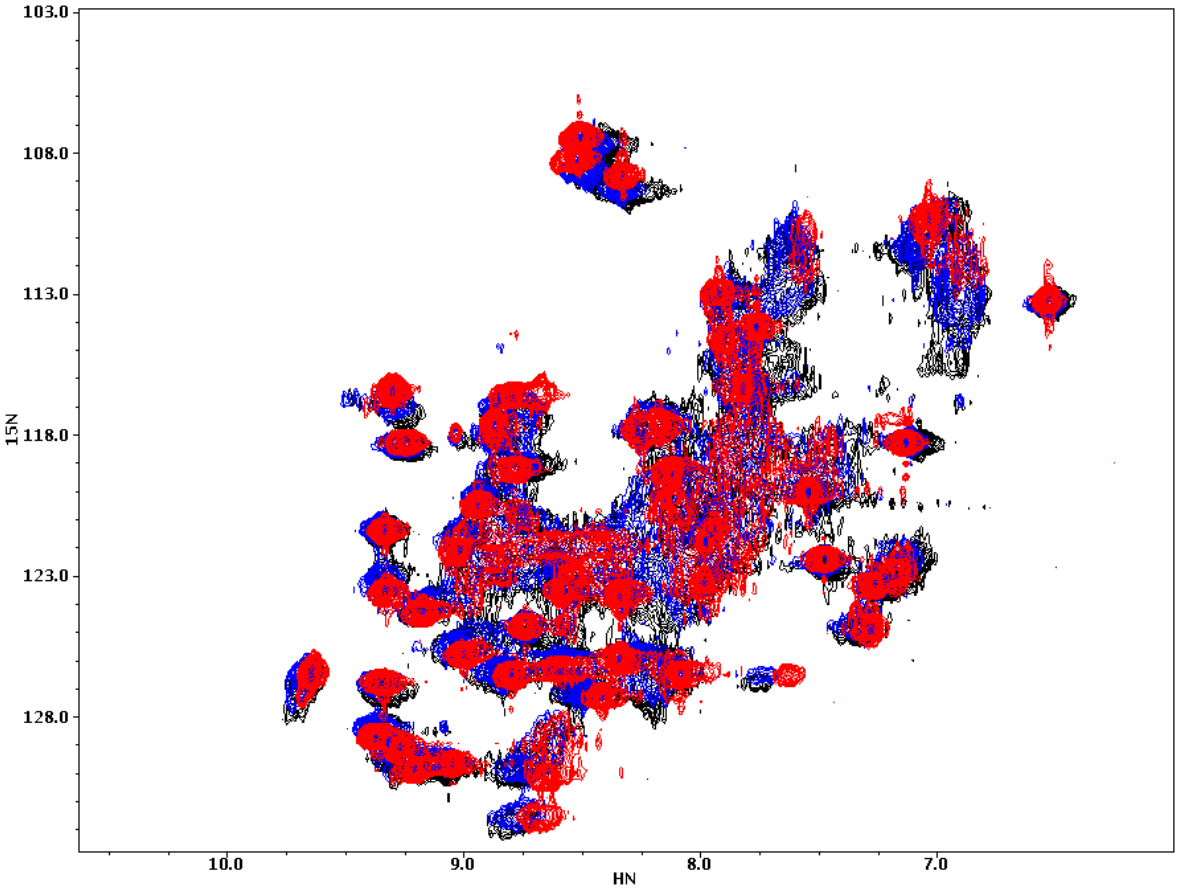
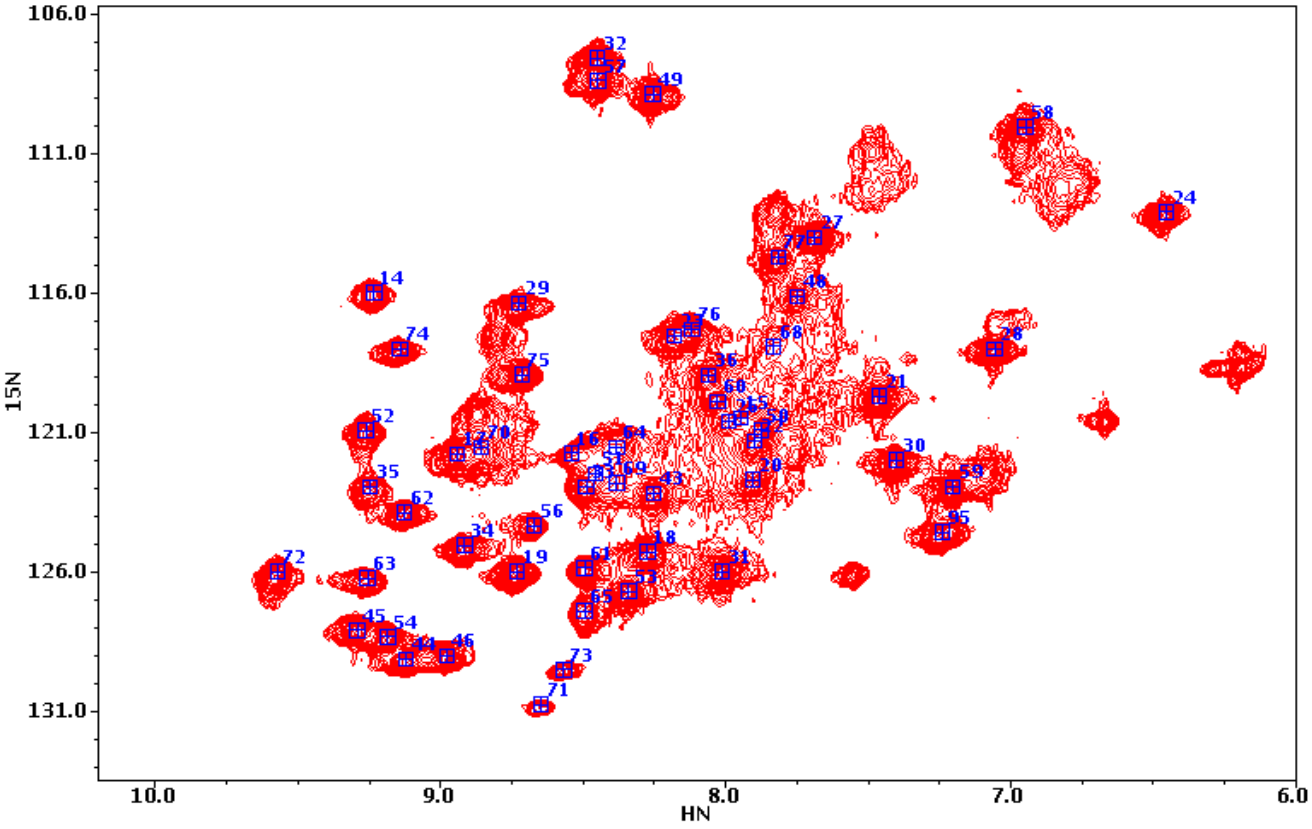


Figure 3.2. Overlay of RPP29 HSQC spectra collected at 25ºC (black), 37ºC (blue), and 55ºC (red). The number of signals observed are very similar at each temperature, however the quality of the spectrum improves with increased temperature. The peaks from the red spectrum are narrowest and the best dispersed.



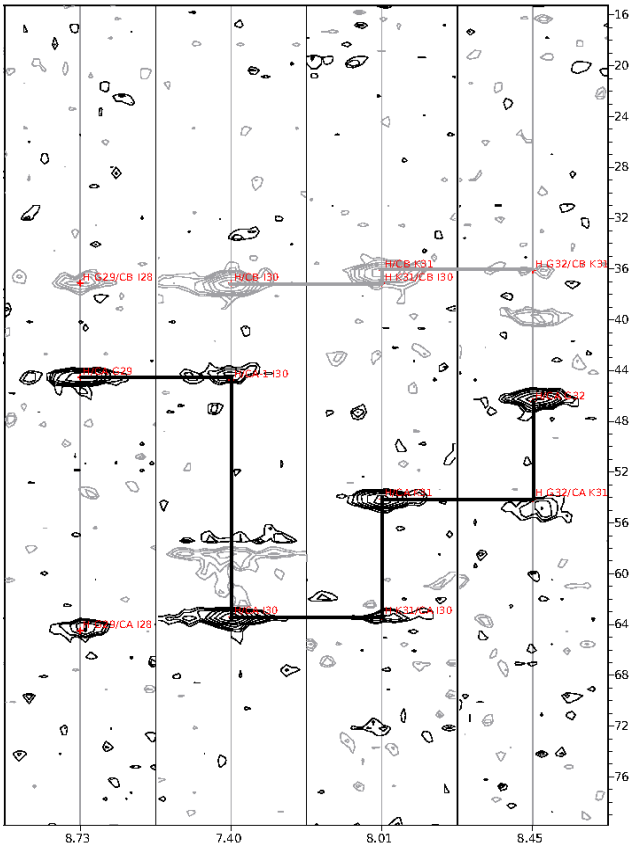


Figure 3.3. (Left) protein amides can be identified through comparisons with nearby chemical shifts. Each HNCACB strip represents the ^{13}C chemical shifts that would appear in the third dimension for a given ^{1}H-^{15}N HSQC peak. (Above) Backbone amide resonance assignments of free RPP29. The numbers represent the numerical position of residues within the protein sequence.

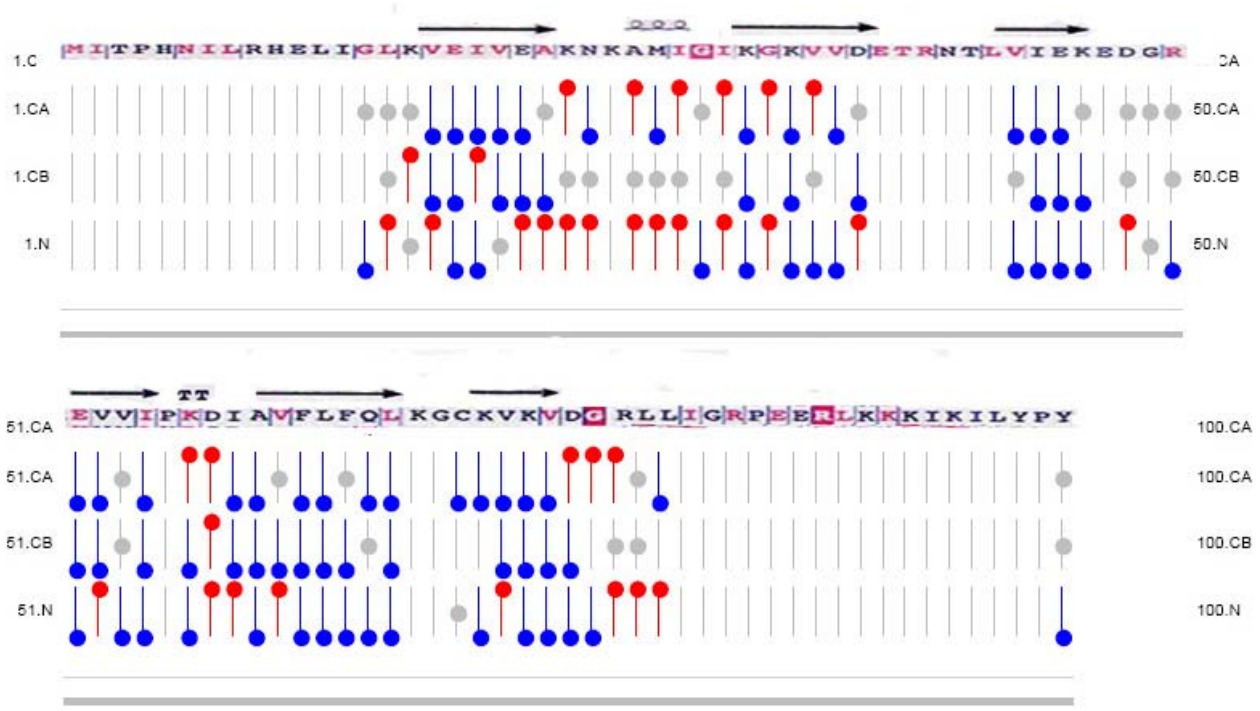


Figure 3.4. A Chemical Shift Index (CSI) displays how the resonance frequencies of nuclei of each residue differs from the random coil value. The difference is represented as a color; red (alpha helix), blue (beta sheet), or gray (random coil). Above the CSI is the *Mja* RPP29 sequence and expected secondary structure from sequence alignment. This information is used to predict secondary structure in solution.

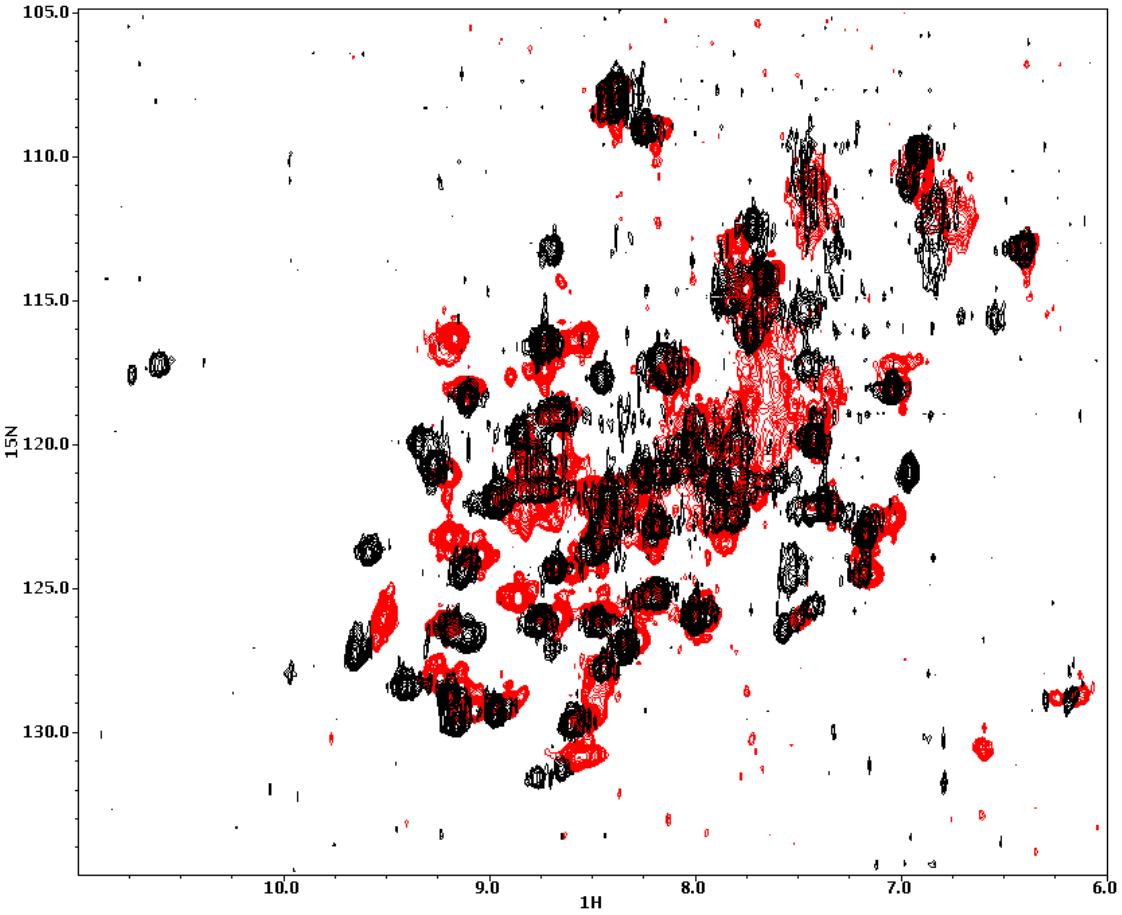


Figure 3.5. Overlay of NMR spectra collected from RPP29 (red) and RPP29 in solution with RPP21 (black). There are a few perturbations between the spectra.

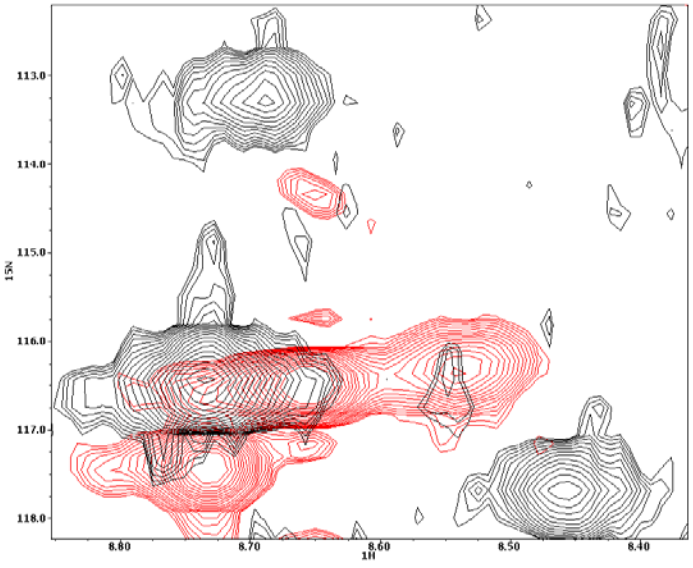


Figure 3.6. Close up view of the RPP29 (red) and RPP29 in solution with RPP21 (black) overlay. Small, but perceptible signals are observed from the RPP29/RPP21 spectrum that overlap with signals from the RPP29 spectrum. This observation indicates that some free RPP29 exists in the RPP29/RPP21 sample.

DISCUSSION

RPP21 deviates from expected behavior

Initial attempts to purify RPP21 were inspired by its expected isoelectric point, pI, of 10.6. Our expectation was that the cationic RPP21 could be easily purified with cation-exchange chromatography; however, successive trials resulted in no binding to the negatively charged SP resin. Similar preparations have been successful [34], however the preparations were performed from a contiguous expression of the RNase P holoenzyme, as opposed to our independent expression of each RPP. Other studies that have required isolated protein, often used RPP21 from type-A organisms, and while the expected pIs are all similar, there are significant deviations between *Mja* and *Pho*/*Pfu* in the locations of non-polar and polar amino acids. No single factor could be identified as causing the observed deviations. Ultimately, purification was managed by cloning a poly-Histidine tag with TEV site to the N-terminus of the protein.

Characterization of RPP29

At the current stage of structural study I am not able to generate a three dimensional model, though several structural motifs have been identified. Our results are mostly consistent with secondary structure predictions made from homology modeling. The series of anti-parallel beta sheets observed in homologous proteins form a beta barrel structure; however there was a peculiar gap in the RPP29 resonance assignments expected for the β2/β3 structural elements. Without further structural restraints, this region remains largely abstruse. The peaks assigned to the β5 strand (residues 68-75), which were either diminished or unobserved in the RPP29/RPP21 spectrum, may be evidence for deviations from the expected structure. Structural studies of *Pho*RPP29 and *Pfu*RPP29 [52, 53] report that the β2 strand and terminal α helices compose the heterodimer interface. This provokes the question; does the *Mja*RPP29-RPP21 interface differ from the RPP29-RPP21 interface observed in archaeal type A

organisms? The eventual resonance assignment of the RPP29-RPP21 complex will provide a more refined picture of RPP29.

The HSQC spectra obtained from {^{15}N, ^{1}H}-RPP29/RPP21 was indicative of incomplete/weak binding. Small but perceptible signals were observed as matching the underlayed free RPP29 spectrum, especially near peaks that have been assigned. This observation may signify a slow exchange between conformations, which would have different shielding effects on the amide, or H-bonding events [46].

The additional peaks seen in the RPP21/RPP29 spectrum are likely due to discursive conformational changes. No doubt the spectra suffered from sample inhomogeneity, but the source cause may be an inherently low association constant, K_A, for RPP21 and RPP29.

Conclusion

Protocols have been successfully developed for production of RPP21 and RPP29 from the archaeon *Mja*. Subsequent NMR experiments were performed on purified RPP29, and the resonance assignment of free RPP29 was completed for over half of the protein. When NMR samples were prepared that included RPP21 and labeled-RPP29, the resulting spectra was only modestly different than free RPP29. This result likely indicates incomplete binding or protein misfolding, no conclusive statement can be made regarding RPP29 structural changes in the presence of RPP21.

ACKNOWLEDGEMENTS

This research was made possible by the frequent guidance of Dr. Mark Foster, Dr. Sri Vidya Oruganti, Dr. Yiren Xu, and other Foster laboratory members. I'm very appreciative of their ideas and knowledge. I also owe thanks to Dr. Venkat Gopalan and his team for research materials and supplemental information about RNase P. Lastly, our research is made possible from resources made available by the Campus Chemical Instrument Center and funding received from the National Institute of Health.

REFERENCES

1. Stark B, Kole R, Bowman E, and Altman S, *Ribonuclease P: an enzyme with an essential RNA component.* PNAS, 1978. **75**: p. 3717-3721.
2. Kruger K, Grabowski P, Zaug A, Sands J, Gottschling D, and Cech T, *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982: p. 147-157.
3. Altman S, *A view of RNase P.* Molecular BioSystems, 2007. **3**(9): p. 604-607.
4. Altman S and Gopalan V, *Ribonuclease P: structure and catalysis*. Third Edition ed. The RNA World, ed. R. Gesteland, T. Cech, and J. Atkins. 2006: Cold Spring Harbor Laboratory Press.
5. Evans D, Marquez S, and Pace N, *RNase P: interface of the RNA and protein worlds.* Trends Biochem Sci, 2006. **31**: p. 333-341.
6. Chamberlain J, Lee Y, Lane W, and Engelke D, *Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive overlap with RNase MRP.* Genes Dev, 1998. **12**: p. 1678-1690.
7. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, and Altman S, *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**: p. 849-857.
8. Pannucci J, Haas E, Hall T, Harris J, and Brown J, *RNase P RNAs from some Archaea are catalytically active.* Proc Natl Acad Sci USA, 1999. **96**: p. 7803-7808.
9. Gopalan V, *RIBONUCLEASE P: Unity and Diversity in a tRNA Processing Ribozyme.* Annu Rev Biochem, 1998. **67**: p. 153-80.
10. Kikovska E, Svärd S, and Kirsebom L, *Eukaryotic RNase P RNA mediates cleavage in the absence of protein.* Proc Natl Acad Sci USA, 2007. **104**: p. 2062-2067.
11. Pulukkunat D and Gopalan V, *Studies on Methanocaldococcus jannaschii RNase P reveal insights into the roles of RNA and protein cofactors in RNase P catalysis.* Nucl Acids Res, 2008. **36**: p. 4172-4180.
12. Gopalan V, *RIBONUCLEASE P: Unity and Diversity in a tRNA Processing Ribozyme.* Proc Natl Acad Sci USA, 2007. **104**: p. 2031-2032.
13. Schedl P and Primakoff P, *Mutants of Escerichia coli thermosensative for the synthesis of transfer RNA.* Proc Natl Acad Sci USA, 1973. **70**(2091-2095).
14. Buck A, Kazantsev A, Dalby A, and Pace N, *Structural Perspcetives on the activation of RNase P RNA by a protein.* Nat Struct Mol Biol, 2005. **12**: p. 958-964.
15. Buck A, Dalby A, Poole A, Kazantsev A, and Pace N, *Protein activation of a ribozyme: the role of bacterial RNase P protein.* EMBO J, 2005. **24**: p. 3360-3368.
16. Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, and Inokuchi H, *A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli.* Proc Natl Acad Sci USA, 1994. **91**: p. 9223-9227.
17. Li Y and Altman S, *Polarity effects in the lactose operon of Escherichia coli.* J Mol Biol, 2004. **339**: p. 31-39.
18. Alifano P, Rivellini F, Piscitelli C, Arraiano C, Bruni C, and Carlomagno M, *Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA.* Genes Dev, 1994. **8**: p. 3021-3031.
19. Altman S, Wesolowski D, Guerrier-Takada C, and Li Y, *RNase P cleaves transient structures in some riboswitches.* Proc Natl Acad Sci USA, 2005. **102**: p. 11284-11289.
20. Kim K and Liu F, *Inhibition of gene expression in human cells using RNase P-derived ribozymes and external guide sequences.* Biochem Biophys Acta, 2007. **1769**: p. 603-612.

21. Eubank T, Gopalan V, Biswas R, Javonovic M, Litovchick A, and Lapidot A, *Inhibition of bacterial RNase P by aminoglycoside-arginine conjugates.* FEBS lett., 2002: p. 107-112.
22. Kazantsev A and Pace N, *Bacterial RNase P: a new view of an ancient enzyme.* Nat Rev Microbiol, 2006. **4**: p. 729-740.
23. Krasilnikov A, Yang X, Pan T, and Mondragón A, *Crystal structure of the specificity domain of ribonuclease P.* Nature, 2003. **421**: p. 760-764.
24. Torres-Larios A, Swinger K, Krasilnikov A, Pan T, and Mondragón A, *Crystal structure of the RNA component of bacterial ribonuclease P.* Nature, 2005. **437**: p. 584-587.
25. Kazantsev A, Krivenko A, Harrington D, Holbrook S, Adams P, and Pace N, *Crystal structure of a bacterial ribonuclease P RNA.* Proc Natl Acad Sci USA, 2005. **102**: p. 13392-13397.
26. Kazantsev A, Krivenko A, Harrington D, Carter R, Holbrook S, Adams P, and Pace N, *High-resolution structure of RNase P protein from Thermotoga maritima.* Proc Natl Acad Sci USA, 2003. **100**: p. 7497-7502.
27. Spitzfaden C, Nicholson N, Jones J, Guth S, Lehr R, Prescott C, Hegg L, and Eggelston D, *The structure of ribonuclease P protein from Staphylococcus aureus reveals a unique binding site for single-stranded RNA.* J Mol Biol, 2000. **295**: p. 105-115.
28. Stams T, Niranjanakumari S, Fierke C, and Christianson D, *Ribonuclease P protein structure: evolutionary origins in the translational apparatus.* Science, 1998. **280**: p. 752-755.
29. Smith D and Pace N, *Multiple magnesium ions in the ribonuclease P reaction mechanism.* Biochemistry, 1993. **32**: p. 5273-5281.
30. Loria A and Pan T, *Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from Bacillus subtillus ribonuclease P.* Biochemistry, 1997. **36**(6317-6325).
31. Nolan J, Burke D, and Pace N, *Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure.* Science, 1993. **261**: p. 762-765.
32. Wegscheid B and Hartmann R, *The precursor tRNA 3'-CCA interaction with Escherichia coli RNase P is essential for catalysis by RNase P in vivo.* RNA, 2006. **12**: p. 2135-2148.
33. Harris J, Haas E, Williams D, Frank D, and Brown J, *New insight into RNase P RNA structure from comparative analysis of the archaeal RNA.* RNA, 2001. **7**: p. 220-232.
34. Andrews A, Hall T, and Brown J, *Characterization of RNase P holoenzymes from Methanococcus jannaschii and Methanothermobacter thermoautotrophicus.* Bio Chem, 2001. **382**(8): p. 1171-1177.
35. Hall T and Brown J, *Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins.* RNA, 2002. **8**: p. 296-306.
36. Kifusa M, Fukuhara H, Hayashi T, and Kimura M, *Protein-protein interactions in the subunits of ribonuclease P in the hyperthermophilic archaeon Pyrococcus horikoshii OT3.* Biosci Biotechnol Biochem, 2005. **69**: p. 1209-1212.
37. Takagi H, Watanbe M, Kakuta Y, Kamachi R, Numata T, Tanaka I, and Kimura M, *Crystal structure of the ribonuclease P protein h1877p from hyperthermophilic archaeon Pyrococcus horikoshii OT3.* Biochem Biophys Res Commun, 2004. **319**: p. 787-794.
38. Boomershine W, McElroy C, Tsai H, Wilson R, Gopalan V, and Foster M, *Structure of Mth11/Mth Rpp29, an essential protein subunit of archaeal and eukaryotic RNase P.* Proc Natl Acad Sci USA, 2003. **100**: p. 15398-15403.
39. Wilson R, Bohlen C, Foster M, and Bell C, *Structure of Pfu Pop5, an archaeal RNase P protein.* Proc Natl Acad Sci USA, 2006. **103**: p. 873-878.
40. Tsai H, Pulukkunat D, Woznick W, and Gopalan V, *Functional reconstitution and characterization of Pyrococcus furiosus RNase P.* Natl Acad Sci USA, 2006. **103**: p. 16147-16152.

41. Jiang T and Altman S, *Protein-protein interactions with subunits of human nuclear RNase P.* Proc Natl Acad Sci USA, 2001. **98**: p. 920-925.
42. Jiang T, Guerrier-Takada C, and Altman S, *Protein-RNA interactions in the subunits of human nuclear RNase P.* RNA, 2001. **7**: p. 937-941.
43. Xu Y, Amero C, Pulukkunat D, Gopalan V, and Foster M, *Solution Structure of an Archaeal RNase P Binary Protein Complex: Formation of the 30-kDa Complex between Pyrococcus furiosus RPP21 and RPP29 Is Accompanied by Coupled Protein Folding and Highlights Critical Features for Protein–Protein and Protein–RNA Interactions.* J Mol Biol, 2009. **392**(1043-1055).
44. Szathmary E, *The origin of the genetic code: amino acids as cofactors in an RNA world.* Trends in Genetics, 1999. **15**: p. 223-229.
45. Brown J, *The Ribonuclease P Database.* Nucl Acids Res, 1999. **27**: p. 314.
46. Cavanagh J, Fairbrother W, Palmer III A, and Skelton N, *Protein NMR Spectroscopy: Principle and Practice*. 1996, San Diego: Academic Press.
47. Delaglio F, Grzasiek S, Vuister G, Zhu G, Pfeifer J, and Bax A, *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**: p. 277-293.
48. Johnson B and Blevins R, J Biomol NMR, 1994. **4**: p. 603-614.
49. Keller R, *The Computer Aided Resonance Assignment Tutorial*. First ed. 2004, Goldau, CH: CANTINA Verlag.
50. Novagen, *pET-15b Vector*.
51. Finkelstein J, Antony E, Hingorani M, and O'Donnell M, *Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy.* Analytical Biochemistry, 2003. **319**: p. 78-87.
52. Amero C, Boomershine W, Xu Y, and Foster M, *Solution structure of Pyrococcus furiosus RPP21, a component of the archaeal RNase P holoenzyme, and interactions with its RPP29 protein parter.* Biochemistry, 2008. **47**: p. 11704-11710.
53. Honda T, Kakuta Y, Kimura M, Saho J, and Kimura K, *Structure of an Archaeal Homolog of the Human Protein Complex Rpp21-Rpp29 That is a Core Component for the Assembly of Active Ribonuclease P.* Journal of Molecular Biology, 2008. **3**: p. 652-662.